昆 虫 学 报 *Acta Entomologica Sinica*, May 2011, 54(5): 548 – 554　　ISSN 0454-6296

斑翅草螽线粒体基因组序列测定与分析

周志军[1,3]，尚　娜[2]，黄　原[3,*]，石福明[1]，韦仕珍[4]

(1. 河北大学生命科学学院，河北保定 071002；2. 河北大学图书馆，河北保定 071002；
3. 陕西师范大学生命科学学院，西安 710062；4. 河池学院化生系，广西宜州 546300)

摘要：已经测定的昆虫线粒体基因组中，直翅目草螽亚科的疑钩额螽 *Ruspolia dubia* 线粒体控制区长度最短，仅 70 bp。为此，本研究采用 L-PCR 结合二次 PCR 扩增策略对另一种草螽亚科昆虫斑翅草螽 *Conocephalus maculates* 线粒体基因组序列进行了测定。序列注释发现：斑翅草螽线粒体基因组序列全长 15 898 bp，A + T 含量为 72.05%，基因排列与典型的节肢动物线粒体基因组一致。全部蛋白质编码基因以典型的 ATN 作为起始密码子，9 个蛋白质编码基因具有完整的终止密码子，其余 4 个以不完整的 T 作为终止信号。除 *trnS*^{AGN}外，其余 21 个 tRNAs 均可折叠形成典型的三叶草结构，依照 Steinberg 等(1997)线粒体特殊 tRNA 结构类型 –9，*trnS*^{AGN}的 DHU 臂形成一个 7 nt 环，反密码子臂则长达 9 bp，含 1 个突起碱基，而不是正常的 5 bp。斑翅草螽与其他直翅目昆虫线粒体基因组的主要区别在于，在 *trnS*^{UCN}和 *nad1*，*nad1* 和 *trnL*^{CUN}基因间各存在一段罕见的、大段的基因间隔序列，长度分别为 78 bp 和 360 bp。其中，位于 *nad1* 和 *trnL*^{CUN}之间的基因间隔序列 N 链可形成一个包含完整起始、终止密码子(ATT/TAA)、编码 103 个氨基酸的未知开放阅读框。同义密码子使用偏好与线粒体基因组编码的 tRNA 反密码子匹配情况无关，但与密码子第 3 位点的碱基组成紧密相关；相对密码子使用频率(relative synonymous codon usage, RSCU)大于 1 的密码子，其第 3 位点全部是 A 或 T。在已经测定的直翅目昆虫线粒体基因组 tRNAs 中，均存在一定数量的碱基错配，且以 G-U 弱配对为主，表明 G-U 配对在线粒体基因组中可能是一种正常的碱基配对形式。本研究测定的斑翅草螽线粒体基因组序列，和先前已经测定的直翅目线粒体基因组序列一起，可以为重建直翅目的进化历史提供数据资源。

关键词：斑翅草螽；线粒体基因组；序列分析；基因间隔序列；相对密码子使用频率
中图分类号：Q966　　**文献标识码**：A　　**文章编号**：0454-6296(2011)05-0548-07

Sequencing and analysis of the mitochondrial genome of *Conocephalus maculatus* (Orthoptera: Conocephalinae)

ZHOU Zhi-Jun[1,3], SHANG Na[2], HUANG Yuan[3,*], SHI Fu-Ming[1], WEI Shi-Zhen[4] (1. College of Life Science, Hebei University, Baoding, Hebei 071002, China; 2. Library, Hebei University, Baoding, Hebei 071002, China; 3. College of Life Science, Shaanxi Normal University, Xi'an 710062, China; 4. Department of Chemistry and Biology, Hechi University, Yizhou, Guangxi 546300, China)

Abstract: The mitochondrial genome (mitogenome) of *Ruspolia dubia* (Orthoptera: Conocephalinae) contains a short control region with 70 bp in length. So, the mitogenome of *Conocephalus maculates* come from Conocephalinae was sequenced by using long-PCR and sub-PCR techniques. The mitogenome of *C. maculates* is 15 898 bp in size, its A + T content is 72.05% and the genome organization follows the ancestral insect gene arrangement. All protein-coding genes (PCGs) start with a typical ATN codon, while nine of the 13 PCGs end with TAA or TAG, and the remainder have incomplete termination codons T. Except for *trnS*^{AGN}, all 21 tRNAs have the typical clover-leaf structures. According to the unusual Type-9 of Steinberg *et al*. (1997), the DHU arm of *trnS*^{AGN} forms a simple 7-nt loop and the anticodon stem has nine base-pairs with a bulged nucleotide in the middle in contrast to the normal five base-pairs. The main differences between *C. maculatus* and other orthopteran mitogenomes was that two novel larger intergenic spacers (78 bp and 360 bp) between *trnS*^{UCN} and *nad1*, *nad1* and *trnL*^{CUN}, respectively. The intergenic spacer between *nad1* and *trnL*^{CUN} has been labeled unidentified open reading frame (UORF) because N strand comprises an open reading frame (103 amino acids) complete with start and termination codons (ATT/TAA). The usage of synonymous codon markedly is correlated with the nucleotide at the 3rd codon

---

基金项目：国家自然科学基金项目(30970346)；教育部高等学校博士学科点专项科研基金项目(20101301120006)；广西高校优秀人才资助计划(RC2007034)

作者简介：周志军，男，1980 年生，山西长治人，博士，讲师，研究方向为直翅目分子进化与系统发育，E-mail：zhijunzhou@163.com
* 通讯作者 Corresponding author, E-mail：yuanh@snnu.edu.cn

收稿日期 Received：2010-12-28；接受日期 Accepted：2011-03-10

position, but has no relation to the anticodon of mitogenome tRNA. Generally, for all codons with their relative synonymous codon usage (RSCU) more than 1, the 3rd codon positions are A or T. Noncanonical matches of G-U base pairs accounts for the overwhelming majority of total unmatched base pairs in the orthopteran mitogenome tRNAs. The results suggest that the G-U may be one cognitive base pairing in mitogenome. The *C. maculates* mitogenome sequences, as well as the previously determined orthopteran mitogenomes, provide sequence source to reconstruct Orthoptera evolutionary history.

Key words: *Conocephalus maculates*; mitochondrial genome; sequence analysis; intergenic spacer; relative synonymous codon usage (RSCU)

昆虫线粒体基因组是一个双链、环状 DNA 分子，长度介于 14 ~ 17 kb，共编码 37 个基因，其中，13 个蛋白质编码基因分别是细胞色素 C 氧化酶 3 个亚基(*cox1*, *cox2*, *cox3*)、细胞色素 *b* 脱氧辅基(*Cytb*)、ATP 合成酶亚基 6 和 8(*atp6* 和 *atp8*)、NADH 脱氢酶亚基 1 ~ 6(*nad1* ~ *nad6*)和 4L(*nad4L*)，2 个 rRNA 基因(*lrRNA* 和 *srRNA*)和 22 个 tRNA 基因(*trnA*, *trnR*, *trnN*, *trnD*, *trnC*, *trnQ*, *trnE*, *trnF*, *trnG*, *trnH*, *trnI*, $trnL^{UUR}$, $trnL^{CUN}$, *trnK*, *trnM*, *trnP*, $trnS^{AGN}$, $trnS^{UCN}$, *trnT*, *trnW*, *trnY*, *trnV*)，组成最小的翻译系统。此外，还包括一个可能与线粒体基因组转录和复制起始调控相关的非编码区，称为控制区(Wolstenholme, 1992; Boore, 1999)。

昆虫线粒体基因组基因间排列通常十分紧密，多数基因间隔序列仅 1 ~ 2 bp，甚至存在碱基重叠，即使是少数较大的基因间隔区，其长度通常也小于 50 bp。尽管如此，在个别种类中也有特大的基因间隔序列被报道，如：膜翅目意大利蜂 *Apis mellifera* 线粒体基因组的 $trnL^{UUR}$ 和 *cox2* 之间的基因间隔序列长度为 193 bp (Crozier and Crozier, 1993)；红光雄蜂 *Bombus ignitus* 线粒体基因组的 $trnS^{UCN}$ 和 *nad1* 之间的基因间隔序列长度为 222 bp (Cha *et al.*, 2007)；半翅目 *Triatoma dimidiata* 线粒体基因组的 $trnS^{UCN}$ 和 *nad1* 之间的基因间隔序列长度为 314 bp，该间隔序列由 2 次重复的完整重复单元(135 bp)与部分重复单元(17 bp 来自重复单元 5′端, 24 bp 来自重复单元中部)组成，该序列具备完整的起始及终止密码子，可能编码一未知基因(Dotson and Beard, 2001)；鞘翅目红锯角萤 *Pyrocoelia rufa* 线粒体基因组的 *nad2* 和 *trnW* 之间的基因间隔序列长度为 1 724 bp，该间隔序列由 12 次重复的完整重复单元(134 bp)与部分重复单元(116 bp 来自重复单元 5′端)组成(Bae *et al.*, 2004)；直翅目 *Xyleus modestus* 的线粒体基因组的 $trnS^{UCN}$ 和 *nad1* 之间的基因间隔序列长度为 259 bp (Sheffield *et al.*, 2010)。这些间隔序列大致分为以下 4 种类型：1)由特定重复单元重复一定次数串联而成(Dotson and Beard, 2001; Bae *et al.*, 2004)；2)具有极高的 A + T 含量，被认为是线粒体基因组的第 2 个复制起始位点(Crozier and Crozier, 1993; Cha *et al.*, 2007)；3)具备完整的起始、终止密码子，编码一定长度氨基酸的未知的开放阅读框(Dotson and Beard, 2001)；4)其他基因间隔序列类型。

Flook 等(1995)测定了第一种直翅目昆虫飞蝗 *Locusta migratoria* 的线粒体基因组全序列。黄原等(2010)对直翅目线粒体基因组研究的一些重要发现进行了综述。目前，GenBank 数据库中共收录直翅目昆虫线粒体基因组全序列数据 34 种，其中，24 种来自蝗亚目，10 种来自螽亚目(http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/)。这些线粒体基因组全序列长度介于 14 971 ~ 17 004 bp 之间，长度差异主要源自控制区长度差异，其中，蝗亚目牛蝗总科 *Physemacris variolosa* 的控制区最长，达 2 277 bp (Sheffield *et al.*, 2010)，我们测定的螽亚目草螽亚科疑钩额螽 *Ruspolia dubia* 拥有目前已报道的最短的线粒体控制区(70 bp) (Zhou *et al.*, 2007)。为此，本文对另一种草螽亚科昆虫斑翅草螽 *Conocephalus maculates* 线粒体基因组全序列进行了测定。

1 材料与方法

1.1 试虫

斑翅草螽 *C. maculates* 标本采集于河北省顺平县(38°83′N, 115°13′E)，标本采集后浸泡于无水酒精中，于 4℃保存，期间更换酒精数次。

1.2 总 DNA 提取

取单头雄性标本的后足股节肌肉，采用常规的酚-氯仿法提取总 DNA。

1.3 PCR 扩增与序列测定

使用2对L-PCR引物(刘念等, 2006)将斑翅草螽的线粒体基因组扩增为2个相互重叠的片段(A和B)。其中, 片段A长度约为8.5 kb, 涵盖线粒体基因组的 *cox1* 至 *cytb* 3′端。使用的引物对为LP03 (5′-CATTTATTTTGATTYTTTGGWCAYCCAGAAGT-3′)和LP04 (5′-AAAATWGCRTAWGCAAATARAAAATATCATTC-3′)。扩增反应程序为: 94℃预变性2 min, 92℃变性10 s, 44℃退火30 s, 前15个循环68℃延伸8 min, 后15个循环每个循环延伸时间增加20 s, 最后68℃延伸7 min。片段B长度约为9.0 kb, 涵盖线粒体基因组的 *cytb* 5′端至 *cox2*。使用的引物对为LPCyt b (5′-WACACCAGTTCATATTDAACCAGAATGATATT-3′)和LPCOII (5′-TGATTAGCTCCACAAATTTCTGAACATTGACC-3′)。扩增反应程序为: 94℃预变性2 min, 94℃变性30 s, 59.6℃退火并延伸15 min, 共30个循环, 最后68℃延伸7 min。

经琼脂糖凝胶电泳切胶纯化后, L-PCR扩增产物被用于二次PCR扩增模板。28条扩增产物长度介于0.6~1.4 kb之间, 第1轮首先采用PCR法双向直接测序, 少数测序结果不理想的片段, 则进行第2轮克隆测序, 克隆使用的载体为pMD 19-T载体(TaKaRa)。测序试剂盒为ABI PRISM BigDye Terminator v3.1, PCR法直接测序引物与二次PCR扩增引物相同。测序仪为ABI PRISM™ 3100-Avant Genetic Analyzer。

1.4 序列组装与注释

斑翅草螽线粒体基因组测序结果拼接与注释使用Staden序列分析软件包进行(Staden *et al.*, 2000)。首先, 使用tRNAscan-SE进行tRNA编码基因注释(Lowe and Eddy, 1997), 少数通过tRNAscan-SE无法识别的tRNA编码基因、蛋白质编码基因、rRNA编码基因则采用Staden软件包中的Spin软件, 通过与近缘种同源序列比对, 结合Blast同源搜索确定。我们使用MEGA 4.0对线粒体基因组各部分的碱基组成和密码子相对使用频率(relative synonymous codon usage, RSCU)进行统计(Kumar *et al.*, 2004)。

2 结果与分析

2.1 基因组基因组成与排列

斑翅草螽线粒体基因组全长15 898 bp, A+T含量72.05% (已提交至GenBank数据库, 序列号为HQ711931)。基因组成和基因间的相对位置与典型的节肢动物线粒体基因组一致, 共编码37个基因, 包括13个蛋白质编码基因、22个tRNA编码基因和2个rRNA编码基因和一段非编码的控制区。除4个蛋白质编码基因(*nad1*, *nad4*, *nad4L* 和 *nad5*)、2个rRNA (*lrRNA* 和 *srRNA*)和8个tRNAs (*trnQ*, *trnC*, *trnY*, *trnF*, *trnH*, *trnP*, *trnL*^{CUN} 和 *trnV*)由N链编码外, 其余23个基因均由J链编码(图1; 表1)。

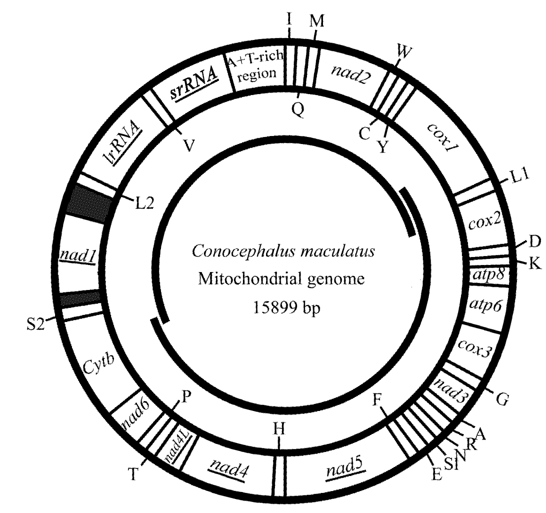


图1 斑翅草螽线粒体基因组组织结构图

Fig. 1 Gene map of the *Conocephalus maculatus* mitogenome

图中深色部分为两个大的基因间隔区, N链编码的蛋白质基因和rRNA基因以下划线标出, tRNA基因于圈内侧以单字母缩写的形式标出。Two novel larger intergenic spacers are marked with dark. PCGs (protein-coding genes) and rRNAs with underline and one-letter symbol of tRNAs within the circle indicates that was coding on N strand.

斑翅草螽线粒体基因组基因间重叠长度为1~8 bp, 共13处, 总长度为37 bp。斑翅草螽线粒体基因组中共包括12处基因间隔序列, 其中, 10处基因间隔序列, 长度为1~3 bp, 总长度为16 bp, 另外两处大的基因间隔序列, 位于 *trnS*^{UCN} 和 *nad1*, 以及 *nad1* 和 *trnL*^{CUN} 之间, 长度分别为78 bp和360 bp。其中, *nad1* 和 *trnL*^{CUN} 之间的基因间隔序列N链可形成一个包含完整起始、终止密码子(ATT/TAA)、编码103个氨基酸的未知开放阅读框。

2.2 蛋白质编码基因

斑翅草螽线粒体基因组中, 13个蛋白质编码基因的起始密码子均为ATN, 分别为ATG 6个(*cox2*, *atp6*, *cox3*, *nad4*, *nad4L* 和 *cytb*), ATT 4个(*atp8*, *nad5*, *nad6* 和 *nad1*), ATA 1个(*nad3*), ATC 2个(*nad2* 和 *cox1*)。9个蛋白质编码基因拥有完整的终止密码子, 分别为TAA 8个(*nad2*, *cox1*, *atp8*, *cox3*, *nad6*, *cytb*, *nad1* 和 *nad4L*)和TAG 1个(*nad3*)。4个蛋白质编码基因以T作为终止信号。

表 1　斑翅草螽线粒体基因组组成

Table 1　Organization of the *Conocephalus maculatus* mitogenome

基因/区域 Gene/Region	位置 Position	长度(bp) Size	方向 Direction	基因间隔 Non	碱基重叠 OL	起始/终止密码子 Start/Termination codon	得分* Score
trnI	1 – 65	65	F				18.74
trnQ	66 – 134	69	R		1		30.86
trnM	134 – 199	66	F				32.70
nad2	200 – 1 231	1 032	F		2	ATC/TAA	
trnW	1 230 – 1 296	67	F	1			33.74
trnC	1 298 – 1 363	66	R	1			23.44
trnY	1 365 – 1 430	66	R		8		18.98
cox1	1 423 – 2 958	1 536	F	2		ATC/TAA	
trnL^{UUR}	2 961 – 3 025	65	F	1			32.26
cox2	3 027 – 3 717	691	F			ATG/T	
trnK	3 718 – 3 787	70	F		1		28.55
trnD	3 787 – 3 854	68	F				35.73
atp8	3 855 – 4 013	159	F		7	ATT/TAA	
atp6	4 007 – 4 679	673	F			ATG/T	
cox3	4 680 – 5 468	789	F	3		ATG/TAA	
trnG	5 472 – 5 537	66	F		3		26.82
nad3	5 535 – 5 891	357	F		2	ATA/TAG	
trnA	5 890 – 5 952	63	F		1		33.01
trnR	5 952 – 6 017	66	F	3			22.34
trnN	6 021 – 6 087	67	F	2			20.01
trnS^{AGN}	6 090 – 6 156	67	F	1			
trnE	6 158 – 6 224	67	F		2		29.82
trnF	6 223 – 6 287	65	R				24.59
nad5	6 288 – 8 019	1 732	R			ATT/T	
trnH	8 020 – 8 086	67	R				15.99
nad4	8 087 – 9 425	1 339	R		7	ATG/T	
nad4L	9 419 – 9 715	297	R	1		ATG/TAA	
trnT	9 717 – 9 780	64	F		1		29.23
trnP	9 780 – 9 844	65	R	1			
nad6	9 846 – 10 367	522	F		1	ATT/TAA	
Cytb	10 367 – 11 506	1 140	F		1	ATG/TAA	
trnS^{UCN}	11 506 – 11 574	69	F	78			51.26
nad1	11 653 – 12 600	948	R	360		ATT/TAA	
trnL^{CUN}	12 961 – 13 025	65	R				33.17
lrRNA	13 026 – 14 330	1 305	R				
trnV	14 331 – 14 400	70	R				39.67
srRNA	14 401 – 15 236	836	R				
控制区 Control region	15 237 – 15 898	662	–	662			

F：表示正向 Forward direction；R：表示反向 Reverse direction；Non：非编码区域 Non-coding region；OL：碱基重叠区域 Overlapping region.

* tRNAscan-SE 预测打分 Score by tRNAscan-SE.

斑翅草螽线粒体基因组蛋白质编码基因不同位点间的碱基组成见表2。13个蛋白质编码基因的平均A+T含量为70.9%，其中，密码子的第3位点含量为83.0%，高于第1位点的64.8%和第2位点的64.7%。

表2 斑翅草螽线粒体基因组不同区域的碱基组成

Table 2 Nucleotide composition in different regions of the *Conocephalus maculatus* mitogenome

	碱基组成 Proportion of nucleotides					碱基数目 Number of nucleotides
	T%	C%	A%	G%	(A+T)%	
全基因组 Whole genome	34.5	16.8	37.6	11.2	72.1	15 898
蛋白质编码基因 Protein-coding genes	40.9	14.9	30.0	14.1	70.9	11 208
密码子第1位点 1st codon position	34.9	14.2	29.8	21.0	64.7	3 736
密码子第2位点 2nd codon position	44.8	20.4	20.0	14.9	64.8	3 736
密码子第3位点 3rd codon position	43.0	10.2	40.3	6.5	83.3	3 736
tRNA基因 tRNA genes	36.0	11.8	37.5	14.8	73.5	1 463
lrRNA	38.9	8.1	36.4	16.6	75.3	1 305
srRNA	38.5	9.8	33.1	18.6	71.6	836
控制区 Control region	37.6	11.8	41.4	9.2	79.0	662

13个线粒体蛋白质编码基因共包括3 736个密码子(不包括4个不完整的终止密码子)，密码子使用情况如表3所示。第3位点是碱基A或T的密码子(3 115个)占密码子总数的83.38%，相对密码子使用频率(RSCU)大于1的密码子，其第3位点全部是A或T。22个与线粒体基因组编码的tRNA反密码子完全匹配的密码子中，仅10个相对密码子使用频率大于1。4种含量最高的氨基酸分别为亮氨酸(16.47%)、丝氨酸(9.02%)、苯丙氨酸(8.77%)和异亮氨酸(8.46%)，占总数的42.72%。

2.3 tRNA和rRNA编码基因

斑翅草螽线粒体基因组的22个tRNA编码基因，20个可以通过tRNAscan-SE识别，仅*trnP*和*trnS*^{AGN}通过同源序列比对的方式加以注释。全部tRNAs的反密码子都与*Drosophila yacuba*一致(Clary and Wolstenholme, 1985)。除*trnS*^{AGN}外，其余21个tRNA均可形成典型的三叶草结构。*trnS*^{AGN}的二级结构则依照Steinberg等(1997)发表的线粒体特殊tRNA结构中类型-9进行预测(Steinberg *et al.*, 1997)，其DHU臂仅形成一个7-nt环，而反密码子臂则长达9 bp，含1个突起碱基(图2)。

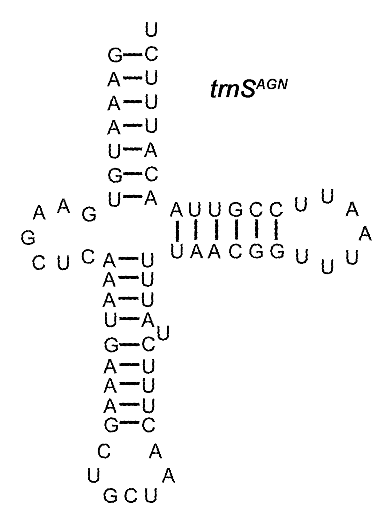


图2 *trnS*^{AGN}的二级结构

Fig. 2 Uncommon secondary structure of *trnS*^{AGN}

斑翅草螽线粒体基因组tRNA中，共存在23处碱基错配现象，其中，19处为G-U形成弱配对，其余4处分别为：A-A，位于*trnM*的反密码子臂；U-U，位于*trnA*的氨基酸接受臂；U-C，分别位于*trnK*和*trnR*的反密码子臂。

斑翅草螽线粒体基因组的*lrRNA*和*srRNA*分别介于*trnL*^{CUN}和*trnV*，*trnV*和线粒体控制区之间。其中，*lrRNA*长度为1 344 bp，A+T含量为75.2%；*srRNA*长度为835 bp，A+T含量为71.7%（表2）。

2.4 控制区

斑翅草螽线粒体基因组的控制区介于*srRNA*和*trnI-trnQ-trnM*之间，长度为662 bp，A+T含量为79%，在线粒体各个组分中仅次于蛋白质编码基因的第3位点(表2)。我们进行进一步序列分析发现：在斑翅草螽线粒体基因组控制区J链存在2个连续的T簇，长度分别为19 bp和10 bp；N链同样存在2个连续的T簇，长度分别为8 bp和9 bp。

表 3　斑翅草螽线粒体基因组蛋白质编码基因的密码子使用

Table 3　Codon usage in protein-encoding genes of the *Conocephalus maculatus* mitogenome

氨基酸 Amino acid	密码子 Codon	数目 n	相对使用频率 RSCU	氨基酸 Amino acid	密码子 Codon	数目 n	相对使用频率 RSCU	氨基酸 Amino acid	密码子 Codon	数目 n	相对使用频率 RSCU
Leu (L)	UUA	385	3.76	Thr (T)	ACU	72	1.36	Trp (W)	UGA	98	1.87
	UUG	50	0.49		ACC	26	0.49		UGG	7	0.13
	CUU	71	0.69		ACA	100	1.89	His(H)	CAU	57	1.36
	CUC	21	0.21		ACG	14	0.26		CAC	27	0.64
	CUA	79	0.77	Met (M)	AUA	180	1.71	Lys (K)	AAA	58	1.40
	CUG	8	0.08		AUG	31	0.29		AAG	25	0.60
Ser (S)	UCU	91	2.17	Val (V)	GUU	92	1.80	Glu (E)	GAA	73	1.80
	UCC	29	0.69		GUC	9	0.18		GAG	8	0.20
	UCA	87	2.07		GUA	87	1.70	Asp(D)	GAU	60	1.62
	UCG	6	0.14		GUG	17	0.33		GAC	14	0.38
	AGU	54	1.29	Ala (A)	GCU	92	1.80	Gln(Q)	CAA	68	1.84
	AGC	15	0.36		GCC	24	0.47		CAG	6	0.16
	AGA	51	1.21		GCA	80	1.57	Arg (R)	CGU	24	1.63
	AGG	3	0.07		GCG	8	0.16		CGC	5	0.34
Phe (F)	UUU	264	1.61	Tyr (Y)	UAU	142	1.65		CGA	24	1.63
	UUC	63	0.39		UAC	30	0.35		CGG	6	0.41
Ile (I)	AUU	265	1.68	Asn (N)	AAU	138	1.63	Cys (C)	UGU	36	1.67
	AUC	51	0.32		AAC	31	0.37		UGC	7	0.33
Gly (G)	GGU	74	1.32	Pro (P)	CCU	76	2.29	Ter.	UAA	8	1.78
	GGC	12	0.21		CCC	16	0.48		UAG	1	0.22
	GGA	96	1.71		CCA	33	0.99				
	GGG	43	0.76		CCG	8	0.24	总计 Total		3 736	

n：密码子使用次数 Number of codon used；RSCU：相对密码子使用频率 Relative synonymous codon usage；下划线标出的密码子与线粒体编码的 tRNA 反密码子完全匹配 The anticodons of the corresponding tRNA were shown with underline；Ter.：终止密码子，不包括 4 个不完全终止密码子 T Termination codon, excluding 4 incomplete termination codons T.

3　讨论

昆虫线粒体基因组基因间排列紧密，基因间隔序列通常只有几个碱基。斑翅草螽线粒体基因组中发现两段大的基因间隔序列(78 bp 和 360 bp)，分别位于 *trnS*^{UCN}和 *nad1*，*nad1* 和 *trnL*^{CUN}之间。类似的情况仅在个别昆虫中被报道，例如意大利蜂、*T. dimidiata*、红锯角萤、红光雄蜂、*X. modestus*（Crozier and Crozier，1993；Dotson and Beard，2001；Bae *et al.*，2004；Cha *et al.*，2007；Sheffield *et al.*，2010）。其中，位于 *nad1* 和 *trnL*^{CUN}之间的基因间隔序列与半翅目 *T. dimidiata* 类似，在其 N 链可形成一个包含完整起始、终止密码子(ATT/TAA)、编码 103 个氨基酸的未知开放阅读框。通过 BLAST 进行序列相似性搜索，在 GenBank 中未发现相似的已知基因序列。

斑翅草螽线粒体基因组中，4 个蛋白质编码基因以不完整的终止密码子 T 作为终止信号。在昆虫线粒体基因组中，以不完整的终止密码子 T 作为终止信号非常普遍，通常认为其可以通过转录后的聚腺苷酸化形成完整的终止密码子（Bibb *et al.*，1981；Ojala，1981）。

在线粒体基因组中，由于密码子简并性，每种氨基酸对应 1～8 种不同的同义密码子。尽管这些氨基酸同时对应多个同义密码子，但是它们总是更偏好于某种特定的密码子。这种偏好与密码子第 3 位点的碱基组成紧密相关，相对密码子使用频率（RSCU）大于 1 的密码子，其第 3 位点全部是 A 或 T。这种偏好可能是由于线粒体基因组 A＋T 含量相对较高所造成的偏倚突变压所引起，即在核苷酸水平上 GC 至 AT 的正向突变大于回复突变（Sueoka，1962）。

斑翅草螽线粒体基因组 tRNA 中，共存在 23 处碱基错配，其中，19 处是 G-U 形成弱配对。已经测定的直翅目昆虫线粒体基因组 tRNA 中，均存在一定数量的碱基错配，且均以 G-U 弱配对为主，约占错配总量的 80% 以上，例如：东方蝼蛄 *Gryllotalpa orientalis*（34 处碱基错配，24 处为 G-U）（Kim *et al*., 2005）；*Anabrus simplex*（21 处碱基错配，18 处为 G-U）（Fenn *et al*., 2007）；疑钩额螽（37 处碱基错配，24 处为 G-U）（Zhou *et al*., 2007）；优雅蝈螽 *Gampsocleis gratiosa*（27 处碱基错配，23 处为 G-U）（Zhou *et al*., 2008）；北方棘螽 *Deracantha onos*（24 处碱基错配，22 处为 G-U）（Zhou *et al*., 2009），这些都表明 G-U 配对在线粒体基因组中很可能是一种正常的配对形式。

本文利用长 PCR 结合二次 PCR 扩增技术测定了斑翅草螽线粒体基因组全序列，主要结论如下：1）斑翅草螽与其他已测定的直翅目昆虫线粒体基因组最大的区别在于，其基因组中存在两段罕见的、大的基因间隔序列。2）在线粒体基因组 tRNA 中，绝大多数的碱基错配是 G-U 形成的弱配对，推测其在线粒体基因组中可能是一种正常的配对形式。由于尚未证实这些大的基因间隔序列是否在全部斑翅草螽个体普遍存在，以及位于 *nad1* 和 *trnL*^{CUN}之间的基因间隔序列是否在转录过程中被转录，G-U 弱配对是否确实是一种正常的配对形式，都有待于转录组水平验证。

参 考 文 献（References）

Bae JS, Kim I, Sohn HD, Jin BR, 2004. The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Mol. Phylogenet. Evol*., 32: 978 – 985.

Bibb MJ, VanEtten RA, Wright CT, Walberg MW, Clayton DA, 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell*, 26: 167 – 180.

Boore JL, 1999. Animal mitochondrial genomes. *Nucleic Acids Research*, 27: 1767 – 1780.

Cha SY, Yoon HJ, Lee EM, Yoon MH, Hwang JS, Jin BR, Han YS, Kim I, 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene*, 392(1 – 2): 206 – 220.

Clary DO, Wolstenholme DR, 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol*., 22: 252 – 271.

Crozier RH, Crozier YC, 1993. The mitochondrial genome of the honeybee *Apis melliferu*: complete sequence and genome organization. *Genetics*, 133: 97 – 117.

Dotson EM, Beard CB, 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol. Biol*., 10(3): 205 – 215.

Fenn JD, Cameron SL, Whiting MF, 2007. The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. *Insect Mol. Biol*., 16(2): 239 – 252.

Flook PK, Rowell CH, Gellissen G, 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitogenome. *J. Mol. Evol*., 41: 928 – 941.

Huang Y, Liu N, Lu HM, 2010. Research progress in mitochondrial genomes of the Orthoptera insects. *Acta Entomologica Sinica*, 53(5): 581 – 586. [黄原, 刘念, 芦慧甍, 2010. 直翅目昆虫线粒体基因组研究进展. 昆虫学报, 53(5): 581 – 586]

Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM, Sohn HD, Jin BR, 2005. The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene*, 353: 155 – 168.

Kumar S, Tamura K, Nei M, 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform*., 5: 150 – 163.

Liu N, Hu J, Huang Y, 2006. Amplification of grasshoppers complete mitochondrial genomes using long PCR. *Chinese Journal of Zoology*, 41(2): 61 – 65. [刘念, 胡婧, 黄原, 2006. 应用长 PCR 扩增蝗虫线粒体全基因组. 动物学杂志, 41(2): 61 – 65]

Lowe TM, Eddy SR, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25: 955 – 964.

Ojala D, Montoya J, Attardi G, 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290: 470 – 474.

Sheffield NC, Hiatt KD, Valentine MC, Song H, Whiting MF, 2010. Mitochondrial genomics in Orthoptera using MOSAS. *Mitochondrial DNA*, 21(3 – 4): 87 – 104.

Staden R, Beal KF, Bonfield JK, 2000. The Staden package, 1998. *Methods Mol. Biol*., 132: 115 – 130.

Steinberg S, Leclerc F, Cedergren R, 1997. Structural rules and conformational compensations in the tRNA L-Form. *J. Mol. Biol*., 266: 269 – 282.

Sueoka N, 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. Natl. Acad. Sci. USA*, 48: 582 – 592.

Wolstenholme DR, 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol*., 141: 173 – 216.

Zhou ZJ, Huang Y, Shi FM, 2007. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A + T-rich region of 70 bp in length. *Genome*, 50(9): 855 – 866.

Zhou ZJ, Huang Y, Shi FM, Ye HY, 2009. The complete mitochondrial genome of *Deracantha onos* (Orthoptera: Bradyporidae). *Mol. Biol. Rep*., 36(1): 7 – 12.

Zhou ZJ, Shi FM, Huang Y, 2008. The complete mitogenome of the Chinese bush cricket, *Gampsocleis gratiosa* (Orthoptera: Tettigonioidea). *J. Genet. Genomics*, 35(6): 341 – 348.

（责任编辑：袁德成）